ARMY RESEARCH LABORATORY

Centrality Measures of Dynamic Social Networks

by Allison Moore

ARL-TN-0513 November 2012

NOTICES

Disclaimers

Army Research Laboratory
Aberdeen Proving Ground, MD 21005

ARL-TN-0513 **November 2012**

Centrality Measures of Dynamic Social Networks

Allison Moore
Computational and Information Sciences Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
November 2012	Final	July 2012

4. TITLE AND SUBTITLE
Centrality Measures of Dynamic Social Networks

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Allison Moore

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: RDRL-CII-C
Aberdeen Proving Ground, MD 21005

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TN-0513

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Modern criminal networks are constantly changing to maintain secrecy, recruit members, and coordinate activities. Attempts to uncover important elements of these networks need to incorporate dynamic trends to provide useful findings and disrupt harmful plans. Our research provides a promising approach whereby analysts can forecast network behavior and stay a step ahead of their adversaries. This report explores the theoretical background of dynamic networks and uses the network measures of degree, closeness, betweenness, and eigenvector centrality over time to conduct network trend analysis. As a case study, I examined the Ali Baba data set that provides messages from a fictitious terrorist cell over a seven-month period. The force-directed Fruchterman-Reingold algorithm was used to visualize the Ali Baba network each month to identify structure, distinguish key players, and understand behavioral roles. Despite the low density of interactions, results revealed the ranking of eigenvector centrality to match the terrorist attack cycle. Several methods for centrality measure prediction are also evaluated, including regression and moving average. Lastly, the results of the removal of a key node from a scale-free criminal network are examined. These examples are an important step in the continuing effort to predict terrorist network behavior.

15. SUBJECT TERMS
Centrality measures, network theory, social network, network disruption

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Sue Kase
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (*Include area code*)
Unclassified	Unclassified	Unclassified	UU	18	(410) 278-4988

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank all of the members of the Tactical Information Fusion Branch, specifically, Dr. Sue Kase and Heather Roy for their continued support. I am grateful to Mark Mittrick and Dr. Kirk Ogaard for their database and data mining expertise. Lastly, a special thanks to Dr. Elizabeth K. Bowman for her mentorship over the past three years.

Student Bio

I have just finished my bachelor's degree from Virginia Tech with a double major in mathematics and statistics. This fall I will begin graduate studies at The University of Georgia to complete my master of science in statistics. I am sponsored by the Student Career Experience Program (SCEP). This is my third summer working at the Tactical Information Fusion Branch (TIFB) on the Social Network Analysis team. My research focuses on statistical applications to networks, including centrality measures, regression analysis, dynamic networks, and time series. After graduation, I plan to become a mathematical statistician for the Department of Defense.

1. Introduction

In today's world, modern criminal networks are constantly changing to maintain secrecy, recruit members, and coordinate activities. Previous research has "focused on analyzing static networks that do not change over time… (when) in real life many networks are inherently dynamic" (*1*). By incorporating dynamic trends one can discover important elements of the network and disrupt harmful plans (*2*). This research provides a potential approach where analysts can compare different centrality values and predict how key players change over time (*3*, *4*).

The purpose of this report is to (1) identify and visualize how a network changes over time, (2) calculate centrality measures of a dynamic network, (3) evaluate prediction methods to forecast network behavior, and (4) examine how a network instantaneously changes when a node is removed. In section 2, I examine the theory behind a visualization algorithm, centrality measures, and prediction methods. In section 3, I apply this theory to a case study of the Ali Baba data set. Lastly, in section 4, I identify continued efforts with this project and future work in the field.

In order to compute the measures, the open source R programming language and environment was used. Over the past year I have been teaching myself R and I wrote more than 400 lines of code for this project. Sample code was taken from Stanford University (*5*) as a starting point for this project. This work was supplemented by the "igraph" package, which implements algorithms and additional call functions for social network analysis (SNA) (*6*). Using this code, I was able to visualize the network and calculate centrality measures.

2. Network Theory

For this report, Fruchterman-Reingold was chosen for its strengths in visualizing large undirected networks using a force directed algorithm. The advantages are its flexibility, simple structure, and interactive nature. Also, the graph establishes edges that are equal in length and reduces the number of intersecting edges. However, the algorithm has a high runtime for extremely large systems. Fruchterman-Reingold assigns forces as if the vertices were electrically charged particles and the edges were springs. Equation 1 represents the energy of the physical system, which is repeated until equilibrium is achieved. The first term is the attraction between connected vertices. The second is the repulsion between pairs of different vertices (*5*, *7*).

$$U(p) = \sum_{\{u,v\}\in E} \frac{1}{3} ||p(u) - p(v)||^3 + \sum_{\{u,v\}\in V} 2 - \ln\ ||p(u) - p(v)|| \tag{1}$$

In a network, individual people are represented as vertices and the relations between them as edges (*4, 8*). This report examines the undirected connections of vectors, or in simpler terms, the association between people (*9*). I investigate the density of a network and take a look into four key centrality measures (*8, 10*) used in network analysis. The density of a network (equation 2) for binary data is defined as the number of edges that exist over the total number of possible edges for the network (*11–13*). Density relates to the speed at which information diffuses between the actors (*11*). Criminal networks typically remain decentralized enough to remain secret, but dense enough to enable coordination (*12*).

$$\frac{E}{V*(V-1)} \tag{2}$$

The "measure of activity" of a network is also known as the degree centrality (equation 3) (*2*). This value determines how many people have direct relationships with an individual (*4, 5, 10*). Degree centrality is a fair approximation of the influence, prominence, or prestige of a node. For simplicity, the more ties a node has (and hence higher degree centrality), the more powerful the person is (*2, 11*).

$$C_d(u) = \frac{\sum r(u,v)}{n-1} \tag{3}$$

Closeness centrality (equation 4) measures the dyad or "number of steps" from each node (u) to all other nodes (v) in the network (*1, 4, 5, 10, 11*). Actors that are close to others are considered more important to the network (*11, 12*). Thus, one can conclude that those with a high closeness centrality are leaders of the network (*2*).

$$C_C(u) = \frac{\sum_{v=1}^{n} d(u,v)}{n-1} \tag{4}$$

Betweenness centrality (equation 5) is a measure of the number of shortest (p_{uv}) going through a specific vertex (w) (*1, 4, 5, 10, 12*). This value determines the gateway between different subgroups and explains the influence over the flow of information (*2*).

$$C_B(w) = \frac{2\sum_{u<v} p_{uv}(w)/p_{uv}}{(n-1)(n-2)} \tag{5}$$

Lastly, eigenvector centrality (equation 6) represents how close an actor is to other actors who are important. An actor can acquire high eigenvector centrality by being connected to a lot of other people or by being connected to others who are highly central (*2, 5, 10*). This centrality is identified by some as the cohesiveness of the group or the connectedness of an individual node (*4*). When the eigenvector centrality value is ranked for nodes within the network, it is called a "node's network importance."

$$x_i = \frac{1}{\lambda}\sum_{j=1}^{n} A_{i,j}x_j \tag{6}$$

Now that I have explained the centrality measures used, I move on to how to predict them. Scientists and analysts agree that even though they are seemingly random, human contact networks are predictable (*1*). If one can forecast network behavior, one can stay a step ahead of our adversaries (*2*). An analyst can use this information to determine when a network is getting too dangerous, decide whether intervening is important, and pick out the key players. Table 1 represents a few of the prediction methods that are investigated.

Table 1. Methods for predicting centrality measures (*1*).

Method	Description
Last	The last node's centrality value
Uniform Moving Average	The average of 'r' past centrality values
Weighted Moving Average	The most recent weighted highest, with decreasing weights
Polynomial Regression	Model of degree three with epsilon term less than 0.2

To analyze the individual methods, the error is computed between the centrality value of interest and the predicted value (equation 7). It is important to remember that there is no single best prediction method for all centrality measures or data sets.

$$Error = \frac{\sum |C(u) - \hat{C}(u)|}{V} \tag{7}$$

When an analyst sees a network that is becoming increasingly dangerous, sometimes network disruption is needed. In network theory, nodes with the highest betweenness are called bridges and those with the highest degree are called hubs. Scale-free networks, where the degree distribution follows a power law, are vulnerable to both bridge and hub removals (*13*). The case study examines several interesting findings regarding prediction and node removal.

3. Case Study

The Ali Baba data set was originally developed in 2003 by the National Security Agency (NSA) to test visualization software. The initial data set contained 752 messages that followed the actions of a fictitious terrorist network centralized in southeast England. The members of the suspected network plan to bomb a water treatment facility as revenge following an outbreak of cholera among Egyptian school children (*14, 15*). Due to the unclassified nature and size of the data set, Ali Baba is commonly used as a testbed for SNA technology (*15*).

Figure 1 shows the Fruchterman-Reingold visualization for the Ali Baba network from May to November. It is important to notice the increasing number of edges and links as well as the visibility of central members.

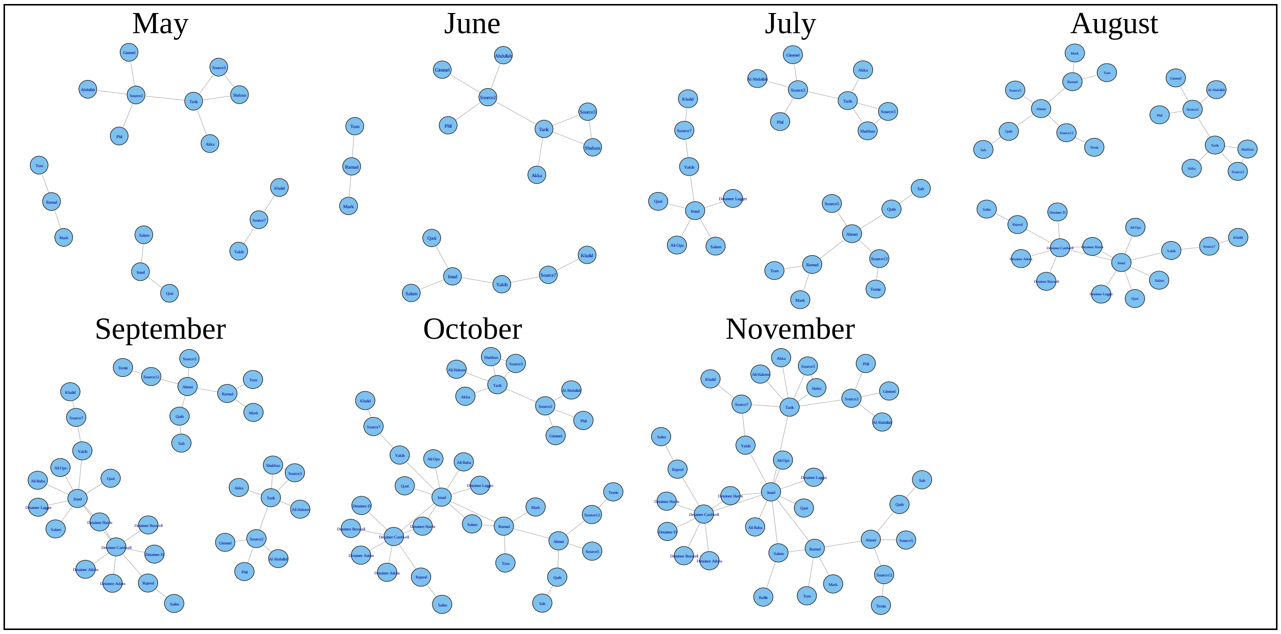

Figure 1. Fruchterman-Reingold network evolution.

First, I look at the density of the entire Ali Baba network from May to November (figure 2). In June, the network is using 11% of all possible ties, and in October, only 6%. This tells us that the members of the network are having as little interaction as possible, which is certainly expected from this sort of criminal network.

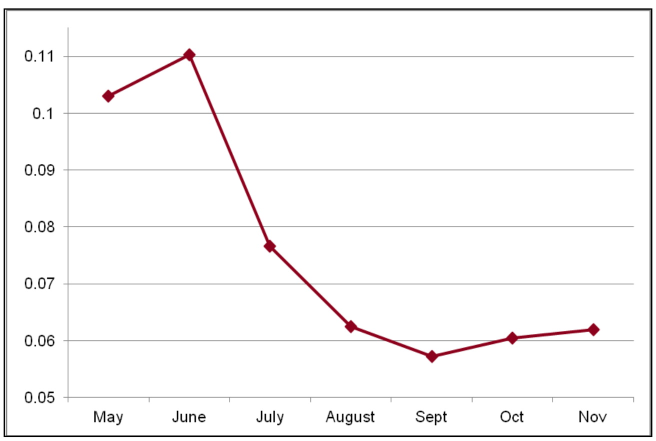

Figure 2. Density of the Ali Baba network.

Now I examine the centrality measures over time of several key members of the Ali Baba network. Looking at figure 3a, there is a large spike in the closeness centrality from October to November, but at this point it is uncertain what that means. Additionally, all individual's closeness centrality measures follow the same general pattern. By just using one centrality measure I was unable to distinguish between the key players.

Next, I examine degree centrality or the "measure of activity." Figure 3b is very helpful in distinguishing between players. In the beginning, Tarik has the most power, but this switches between June and July. As the network grows, Imad takes over as the most powerful individual. This piece of information is important when it comes to network disruption and node removal. Ali Ops and Phil are maintaining their degree measures and current power positions during the last few months.

Betweenness centrality tells us about the flow of information. Figure 3c shows that Imad and Tarik have the greatest influence over the flow of information in November. But the most interesting piece is Phil, who has zero betweenness centrality. If one were trying to disrupt the chain of information, one would not want Phil to be involved because there would be no effect on the network.

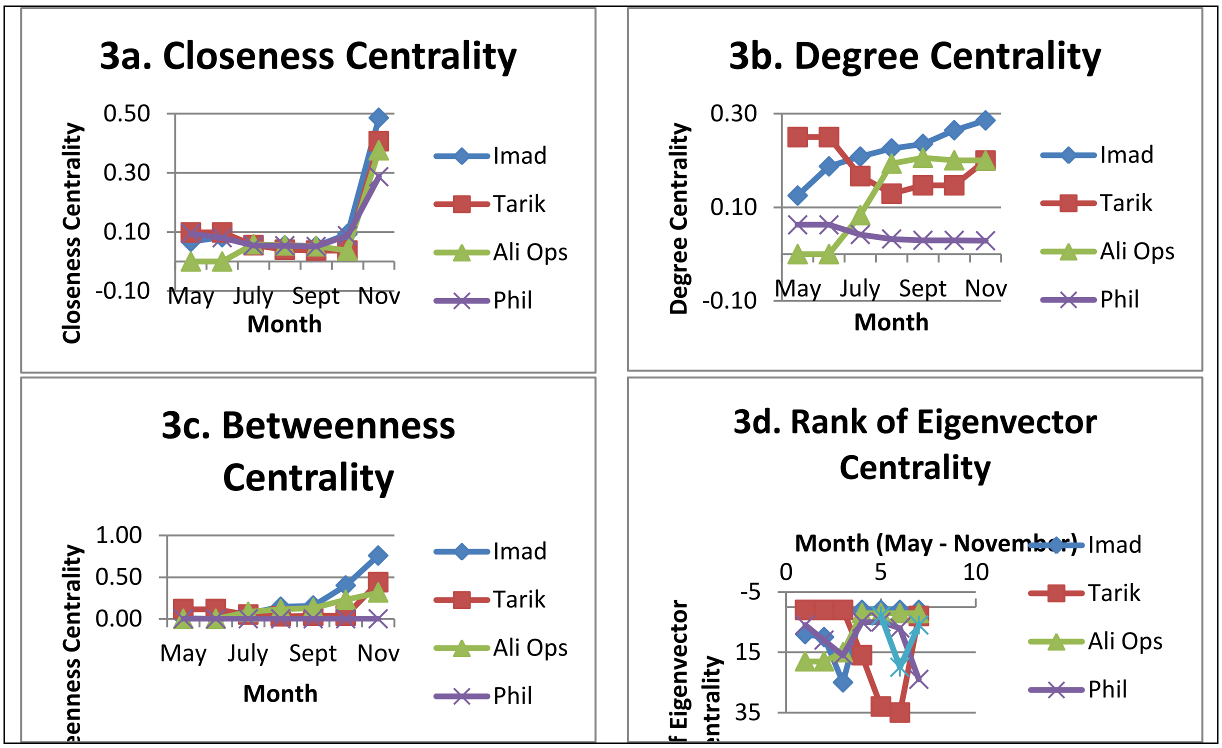

Figure 3. Centrality measures of Ali Baba key members.

The last centrality measure of interest is Eigenvector centrality or a "node's network importance." In red, Tarik is the most important individual from May to July. During July and August there is a big shift in the network eigenvector centralities. Tarik moves from number one to 16th and in his place are Imad and Ali Ops with number one and two spots, respectively. It is interesting to note that in September, Ali Baba appears as the number three person in the network. Ali Baba has very few links in the network so it is exciting to see him stand out with this metric.

In the previous section, the motivation and techniques used for prediction of dynamic social networks were explained. Figure 4 illustrates an example of these prediction methods for degree centrality. Figure 4a compares the four techniques and the original centrality values for Imad. By computing the error, I determined that a third degree polynomial is the best fit and uniform moving average is the worst. Figure 4b applies the polynomial regression to all key players to predict for the month of December. It appears that Imad and Tarik will continue to gain power, while Ali Ops will significantly decrease.

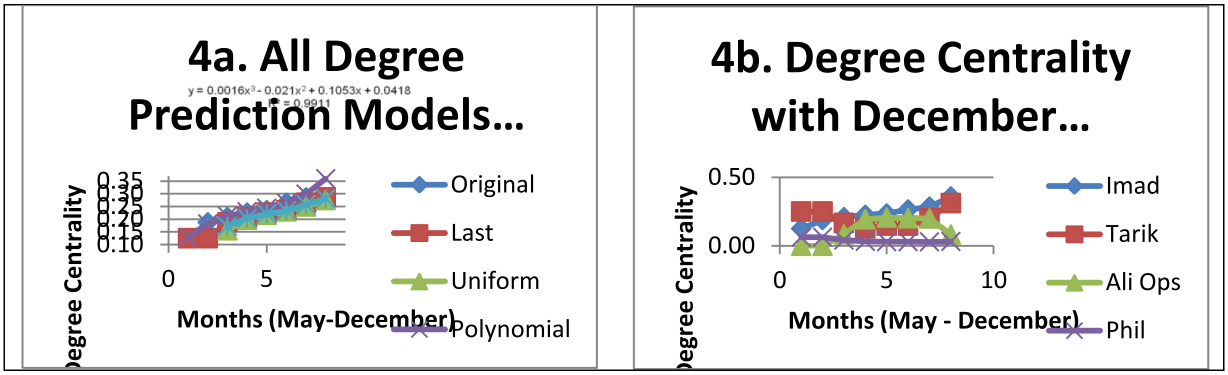

Figure 4. Prediction for degree centrality.

The last thing that I examine is the effect of node removal on the Ali Baba network. Throughout my investigation of centrality measures, Imad stands out as the most important individual. As a bridge and a hub in this scale-free network, I remove Imad to see the instantaneous effect on the network as a whole and its centrality measures. Figure 5 shows the removal of Imad in maroon, which also leads to the removal of direct links and the loss of members whose on ties were with Imad.

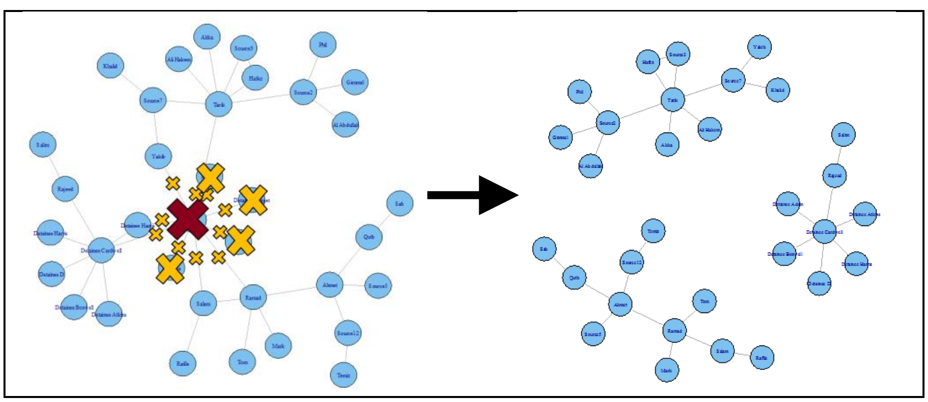

Figure 5. Ali Baba network with node removal.

The centrality measure results of a node removal are shown in figure 6. When Imad is removed, the closeness and betweenness centrality drop to almost nothing. Also notice that Ali Ops was one of the players temporarily removed since his only tie was with Imad. The degree centrality of Tarik and Phil is not significantly changed, so one can assume that they maintain most of their ties. Since Ali Ops is removed, he has zero degree centrality. Lastly, eigenvector centrality (figure 6d) shows that after Imad is removed, Tarik steps up as the number one individual in the network. Phil also becomes more important, by becoming the eighth ranking individual.

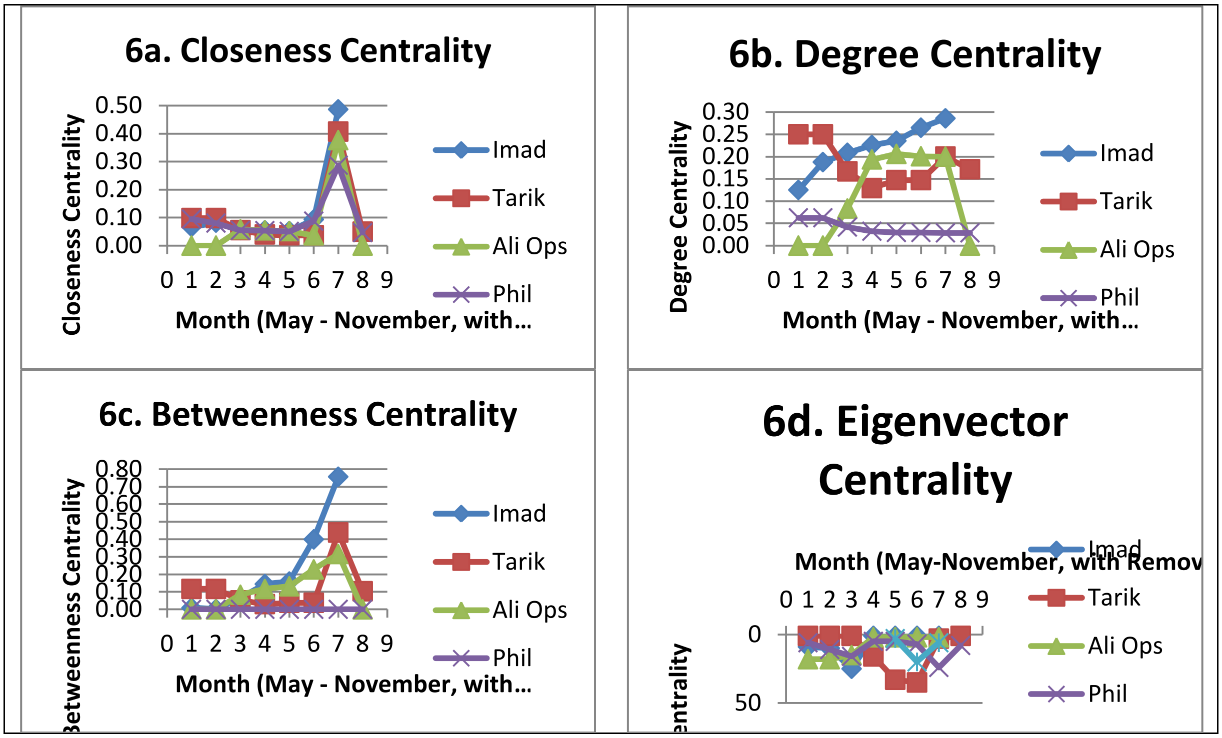

Figure 6. Result of Imad node removal on centrality measures.

4. Conclusions

Throughout this summer project, I continued efforts to predict seemingly random criminal network behavior by viewing them as dynamic systems. I used the Fruchterman-Reingold algorithm to identify structure, distinguish key players, and understand behavioral roles. I conducted a network trend analysis by looking into degree, closeness, betweenness, and eigenvector centrality. I evaluated several methods of centrality prediction including polynomial regression and moving average. Lastly, I looked into the immediate result of removing a key node from a scale free terrorist network.

Additional prediction research can be done using time series analysis and quality control charts to further understand centrality measures of dynamic networks. Evaluation of other metrics such as embeddedness, reachability, and assortivity (*13*) can also be included in this analysis. There is a need for in-depth research on node removal and how networks adapt to the disruption. In the future, I plan to incorporate these additional topics with a secondary case study of the more complex Enron data set (*16*).

5. References

1. Kim, H.; Tang, J.; Anderson, R.; Mascolo, C. Centrality Prediction in Dynamic Human Contact Networks. *Computer Networks* **2012**, *56* (3), 983–96.
2. Memon, N.; Larsen, H. L.; Hicks, D. L.; Harkiolakis, N. Detecting Hidden Hierarchy in Terrorist Networks: Some Case Studies. *Lecture Notes in Computer Science* **2008**, *5075*. Intelligence and Security Informatics, 477–89.
3. Bowman, E. K. *Cultural Situational Awareness for Counter-Insurgency Operations*. U.S. Army Research Laboratory: Aberdeen Proving Ground, 2012.
4. Diesner, J.; Carley, K.M. Exploration of Communication Networks from the Enron Email Corpus. Accepted for *Proc. of Workshop on Link Analysis, Counterterrorism and Security at SIAM International Conference on Data Mining 2005*. Newport Beach, CA, April 21–23, 2005.
5. McFarland, D.; Messing, S.; Nowak, M.; Westwood, S. J. Social Network Analysis Labs in R. Stanford University, 2010.
6. Csardi G., Nepusz T. The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. http://igraph.sf.net.
7. Beyer, D. Force-Directed Graph Layout. *CCVisu 3.0 Introduction and Reference Manual*. University of Passau, 3 Mar. 2010. Web. 25 June 2012. http://ccvisu.sosy-lab.org/manual/.
8. Brandes, U. Social Network Analysis and Visualization. *Signal Processing Magazine* **2008**, *25* (6), 147–51.
9. Yang, C. C. Knowledge Discovery and Information Visualization for Terrorist Social Networks. *Studies in Computational Intelligence* **2008**, *135*, 45–64.
10. Landherr, A.; Friedl, B.; Heidemann, J. A Critical Review of Centrality Measures in Social Networks. *Business & Information Systems Engineering* **2010**, *2* (6), 371–85.
11. Hanneman, R. A.; Riddle, M. *Introduction to Social Network Methods*. Riverside, CA. University of California, 2005. Web. 4 June 2012. <http://faculty.ucr.edu/~hanneman/>.
12. Moody, J. Seminar on Social Networks. *Duke University: Department of Sociology*. PowerPoint. 6 June 2012. http://www.soc.duke.edu/~jmoody77/s884/ syllabus_05.html.
13. Xu, J. The Topology of Dark Networks. *Communications of the ACM* **1 Oct 2008**, *51* (10), 58.
14. Jaworowski, M.; Pavlak, S. Ali Baba Scenario 1 Description. 2003, 2 June 2012.

15. Mittrick, M.; Roy, H.; Kase, S.; Bowman, E. *Refinement of the Ali Baba Data Set*; ARL-TN-0476; U.S. Army Research Laboratory: Aberdeen Proving Ground, 2012.

16. Cohen, W. Enron Email Dataset. Carnegie Mellon University: School of Computer Science, 21 Aug. 2009. Web. 05 June 2012. http://www.cs.cmu.edu/~enron/.

NO. OF COPIES	ORGANIZATION
1 ELEC	ADMNSTR DEFNS TECHL INFO CTR ATTN DTIC OCP 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218
1	ARMY G2 ATTN D WALSH 4812 EXECUTIVE DR GAINSVILLE VA 20155
1	US ARMY CERDEC I2WD ATTN RDER IWP D PORTER 6003 COMBAT DR ABERDEEN PROVING GROUND MD 21005
1	MULTISOURCE INFORMATION FUSION ATTN RESEARCH PROFESSOR (EMERITUS) J LLINAS 315 BELL HALL NORTH CAMPUS STATE UNIVERSITY OF NEW YORK AT BUFFALO BUFFALO NY 14260
1	U.S. ARMY RSRCH LAB ATTN RDRL CII C E K BOWMAN BLDG 390A RM 112 ABERDEEN PROVING GROUND MD 21005
10	US ARMY RSRCH LAB ATTN RDRL CII C M MITTRICK ATTN RDRL CII C S KASE ATTN RDRL CII C A NEIDERER ATTN RDRL CII C B BODT ATTN RDRL CII C C HANSEN ATTN RDRL CII C E HEILMAN ATTN RDRL CII C F S BRUNDICK ATTN RDRL CII C H ROY ATTN RDRL CII C J DUMER ATTN RDRL CII C J RICHARDSON ABERDEEN PROVING GROUND MD 21005
1	US ARMY RSRCH LAB ATTN RDRL CII C M THOMAS BLDG 321 RM 1B ABERDEEN PROVING GROUND MD 21005
1	US ARMY RSRCH LAB ATTN RDRL CII C T HANRATTY BLDG 321 RM 4A ABERDEEN PROVING GROUND MD 21005
8	US ARMY RSRCH LAB ATTN IMAL HRA MAIL & RECORDS MGMT ATTN RDRL CII A S H YOUNG ATTN RDRL CII B BROOME ATTN RDRL CII B L TOKARCIK ATTN RDRL CII B R WINKLER ATTN RDRL CII T V M HOLLAND ATTN RDRL CIO LL TECHL LIB ATTN RDRL CIO LT TECHL PUB ADELPHI MD 20783-1197